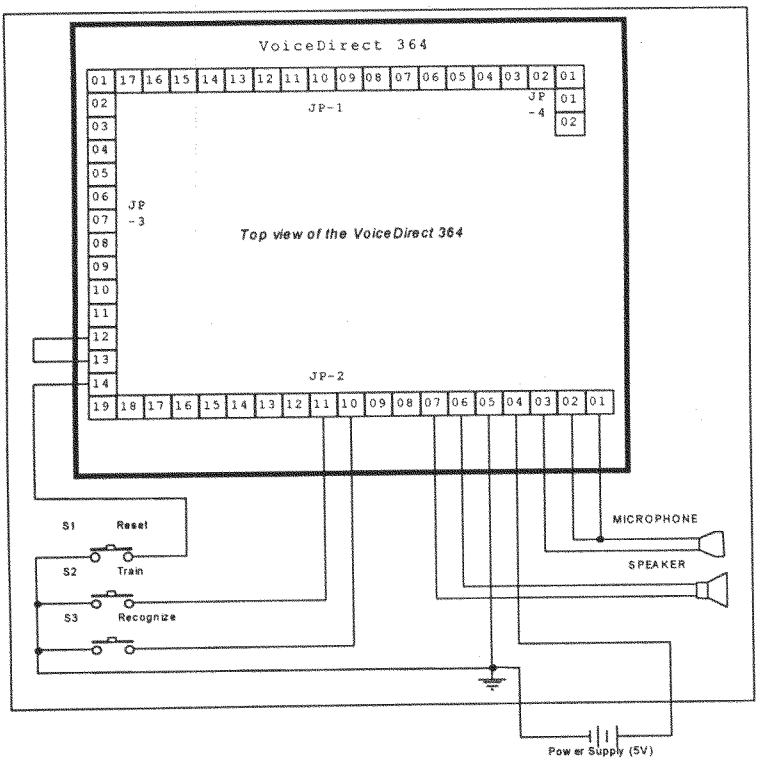
MODE SELECTION

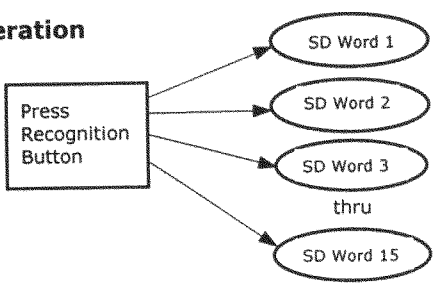
There are 2 main configuration modes for the module that can be selected depending on how voice recognition will interact with your application. These are Speaker-Dependent (SD) mode and Continuous Listening (CL) Mode. There are 2 types of CL modes: single word and multi-word. Following are schematics depicting how to configure the module for each mode. Operating instructions can be found in sections 4, 5 and 6 of this manual.

1. SD Configuration

Figure 1



SD Operation

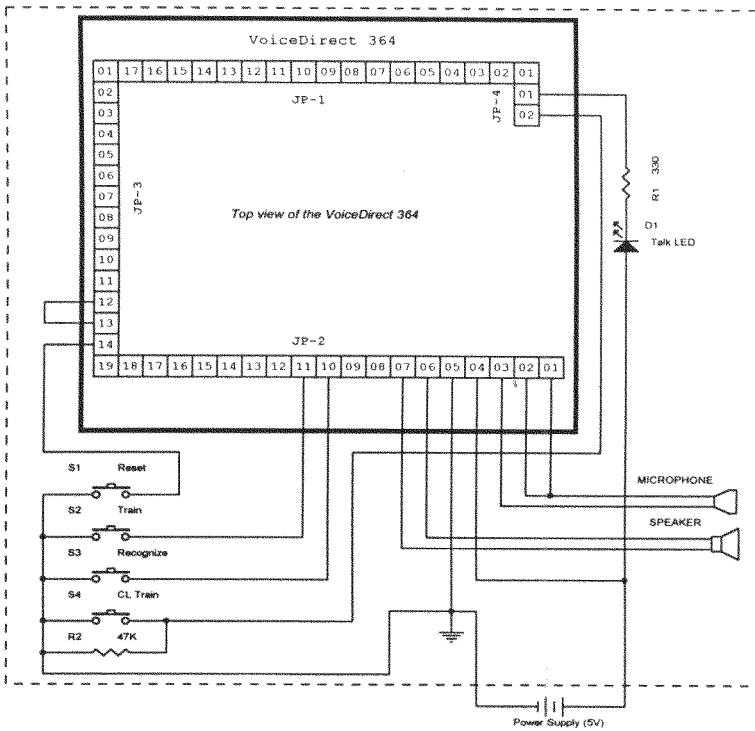


2. Continuous Listening (CL) Mode Configuration

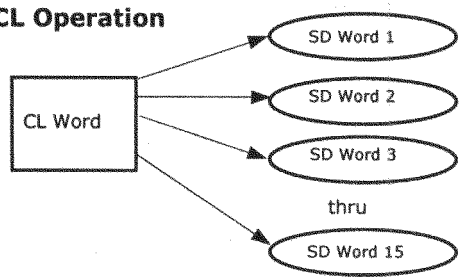
There are 2 separate configurations of CL mode:

Single Word CL Configuration

Figure 2

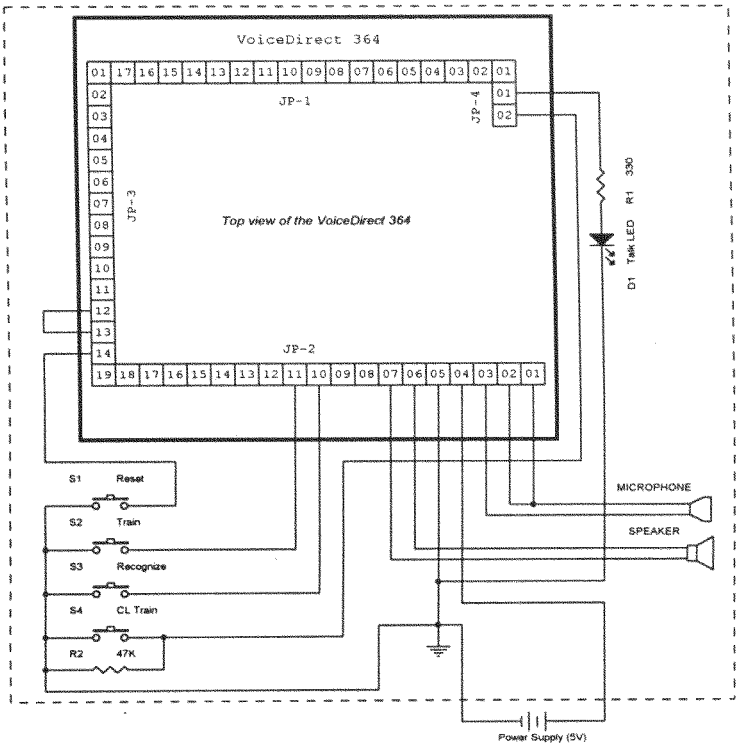


Single Word CL Operation

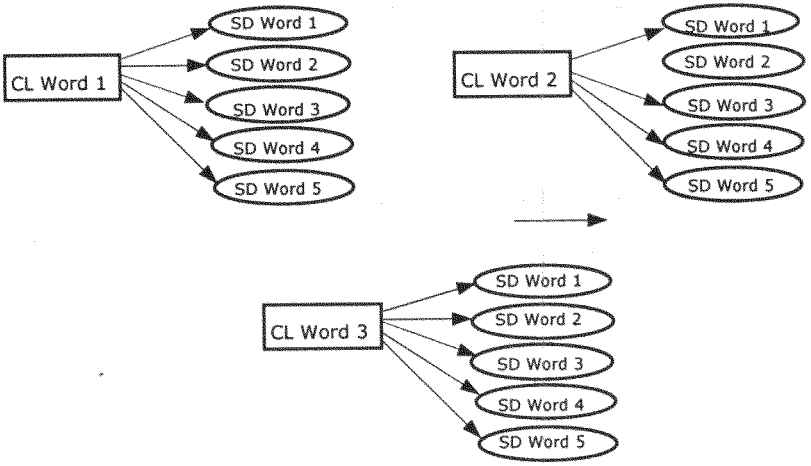


Multi-word CL Configuration

Figure 3



Multi-word CL Operation



Pin #	Name	Description	I/O
JP1:			
01 ~ 16	-	Not Used – Leave open.	I
JP2:			
01	MIC IN	Microphone Input Connection	I
02	MIC BIAS	Microphone Bias (Electret microphone)	O
03	MIC RETURN	Analog Ground. For noise reasons, analog and digital grounds should connect together only at the Voice Direct 364	-
04	+5V	5 Volt (+) power supply connection	I
05	GND	Ground connection	I
06	SPEAKER (+)	Speaker connection (positive)	O
07	SPEAKER (-)	Speaker connection (internally connected to ground)	O
08	DACOUT	Analog output (unbuffered)	O
09	-MHS (P0.0)	Stand Alone Mode: Not Used Slave Mode: Serial Interface Master Handshake (MHS)	- I
10	-RECOG (P0.1)	Stand Alone Mode: Recognition Sensitivity (see below) Slave Mode: Not Used	I -
11	-TRAIN (P0.2)	Stand Alone Mode: Training Sensitivity (see below) Slave Mode: Not Used	I -
12	OUT1 (P1.0)	Stand Alone Mode: Output 1 or 9 (Active high) Slave Mode: Talk	O O
13	OUT2 (P1.1)	Stand Alone Mode: Output 2 or 10 (Active high) Slave Mode: Configurable I/O #0	O I/O
14	OUT3 (P1.2)	Stand Alone Mode: Output 3 or 11 (Active high) Slave Mode: Configurable I/O #1	O I/O
15	OUT4 (P1.3)	Stand Alone Mode: Output 4 or 12 (Active high) Slave Mode: Configurable I/O #2	O I/O
16	OUT5 (P1.4)	Stand Alone Mode: Output 5 or 13 (Active high) Slave Mode: Configurable I/O #3	O I/O
17	OUT6 (P1.5)	Stand Alone Mode: Output 6 or 14 (Active high) Slave Mode: Configurable I/O #4	O I/O
18	OUT7 / -SHS (P1.6)	Stand Alone Mode: Output 7 or 15 (Active high) Slave Mode: Serial Interface Slave Handshake (SHS)	O I/O
19	OUT8 / -DATA (P1.7)	Stand Alone Mode: Output 8 (Active high) Slave Mode: Serial Interface Data (DATA)	O I/O
JP3:			
01 ~ 09	-	Not Used – Leave open.	-
10	ERROR / -PROCESS (P0.3)	Stand Alone Mode: Error LED Slave Mode: Low when processing a command	O O
11	GND	Ground	I
12	GND	Ground	I
13	MODE	Slave/Stand-alone mode selection	I
14	-RESET	Reset (Active low)	I
JP4:			
01	-TALK / CL CFG (P0.6)	Stand Alone Mode: TALK LED, Single-/Multi-CL config. port Slave Mode: Not Used	I/O -
02	-CL TRAIN (P0.7)	Stand Alone Mode: CL Training, CL performance config. port Slave Mode: Not Used	I -

Training and Recognition Selectivity

The –TRAIN, -CL TRAIN and –RECOG pins control the selectivity and activation of the VOICE DIRECT™ 364. The training and recognition selectivity levels are set when the VOICE DIRECT™ 364 is first powered up (or reset) according to the following table:

Pin Configuration	Selectivity
-TRAIN open circuit	Relaxed SD Training – Easier to train, accepts more similar sounding words (fewer rejections).
-TRAIN pulled to GND with a 100K resistor	Strict SD Training – Harder to train, rejects more similar sounding words – may result in better recognition accuracy (fewer substitutions).
-CL TRAIN pulled to GND with a 47K resistor	Relaxed CL Training – Easier to train, accepts more similar sounding words (fewer rejections).
-CL TRAIN pulled to GND with a 680K Resistor	Strict CL Training – Harder to train, rejects more similar sounding words – may result in better recognition accuracy (fewer substitutions).
-RECOG open circuit	Relaxed SD and CL Recognition – Recognizes more words, may substitute the wrong words (fewer rejections).
-RECOG pulled to GND with a 100K resistor	Strict SD and CL Recognition – Recognizes fewer words, may reject trained words (fewer substitutions).

Some applications of Continuous Listening require maximum accuracy with the fewest false triggers. This configuration is appropriate for an application such as a light switch, where a false trigger may be more than a minor inconvenience. Further, the CL recognizer in such an application hears a great many phrases that are *not* trigger phrases, so there are many opportunities for false triggers and a high level of accuracy is desired. A consequence of this high accuracy is relatively slow responsiveness.

Other applications – games, for example – can accept occasional false triggers, but need fast response. In these applications the relative frequency of trigger phrases may be quite high, so false triggering accuracy is less of a concern.

The tradeoff between CL accuracy and responsiveness is controlled by selection of R2. For slow CL performance, select a 680 K-Ohm resistor for R2. For fast CL performance, select a 47 K-Ohm resistor.

Module Pin Characteristics

Absolute Maximum Ratings

Any pin to GND	-0.1V to +6.5V
Operating temperature(T_O)	-20°C to +70°C
Power dissipation	1 W
Operating Conditions	-20°C to +70°C;
	V_{DD}=4.5 - 5.25V
	V_{SS}=0V

WARNING: *Stressing the Voice Direct™ 364 beyond the "Absolute Maximum Ratings" may cause permanent damage. These are stress ratings only. Operation beyond the "Operating Conditions" is not recommended and extended exposure beyond the "Operating Conditions" may affect device*

DC Characteristics

(T_O = -20°C to +70°C, Vdd = 5V)

SYMBOL	PARAMETER	MIN	TYP	MAX	UNITS	TEST CONDITIONS
V_{IL}	Input Low Voltage ♦ MODE, I/O pins,	-0.1		0.75	V	
V_{IH} (Vdd>4.5)	Input High Voltage ♦ MODE, I/O pins,	3.0		Vdd+ 0.3	V	
V_{OL}	Output Low Voltage		0.3	0.1*Vdd	V	I_{OL}= 2 mA
V_{OH}	Output High Voltage	0.8* Vdd	0.9* Vdd		V	I_{OL}= -2 mA
I_{IL}	Logical0 Input Current		<1	10	uA	$V_{ss}<V_{pin}<V_{dd}$
I_{DD1}	Supply Current, Operating		37	100	mA	Hi-Z Outputs
I_{DD2}	Supply Current, Quiescent		4	6	mA	Hi-Z Outputs
Rpu	Pull-up resistance, P0.0-P1.7	5, 80, Hi-Z	6.5, 200, Hi-Z		K-ohm	Selected with software

5. Training VoiceDirect™ 364

SD Mode

The module should be configured for SD Mode (Figure 1). When the –TRAIN pin is pulled to GND for at least 100mS (such as by pushing a momentary switch, labeled in the schematics as "TRAIN"), training will begin. VOICE DIRECT™ 364 will prompt "Say word *x*" (where *x* corresponds to the word to be trained).

A trained word or phrase must be shorter than 2.5 seconds and may not contain silences longer than 0.5 seconds. For example, the name "John Smith" would be an acceptable phrase as long as the two words are not separated by a large pause. Training terminates when no word is spoken in response to a prompt, when either switch is pressed a second time during training, when three errors have occurred during training, or after all words have been trained.

Pressing the TRAIN switch at a later time resumes training. New words are added to the end of the set already recorded. New words may be added to the set at any time, up to a maximum of 15 words. Individual words from the set may not be deleted or overwritten, but the entire set can be erased (see "Erasing Templates", section 7). The user says the first word to be trained, then VOICE DIRECT™ 364 prompts again with "Repeat". The user repeats the word, and VOICE DIRECT™ 364 will return "Accepted" if the word has been successfully trained, otherwise it will say the cause of the training error. If an error occurs during training, then the error will be spoken "Spoke too soon", "Please talk louder", etc., and the ERROR line will pulse high for 1 second. The user will get three attempts to train each word before VOICE DIRECT™ 364 exits training mode, says "Training complete", and pulses the ERROR line high for 1S.

The user can exit training at any time by pressing the –TRAIN or the –RECOG button, by not responding to a "Say word *x*" or "Repeat" prompt, or when all 15 words have been trained.

Example (SD):

Action:	**Press TRAIN switch**	***Start training***
Voice Direct 364:	"Say word one"	*Training first SD word*
User:	"Start"	
Voice Direct 364:	"Repeat"	
User:	"Start"	
Voice Direct 364:	"Say word two"	*Training second SD word*
User:	"Shutdown"	
Voice Direct 364:	"Repeat"	
User:	"Shutdown"	
Voice Direct 364:	"Say word three"	*Training third SD word*
User:	"Lights"	
Voice Direct 364:	"Repeat"	
User:	"Lights"	
Action:	**Press TRAIN switch**	***Stop training***

Note: Voice Direct 364 will continue to prompt for new words to train until all 15 memory locations have been filled. To stop training, press the TRAIN switch again.

Single Word CL Mode

The module should be configured for Single Word CL Mode (Figure 2). In this mode, separate buttons are used to train SD words (TRAIN switch) and the CL word (CL TRAIN switch). After training the CL word, up to 15 SD commands can be trained. Another difference from SD mode is that while training SD words, the TRAIN switch must be pressed to train each word.

Example (Single Word CL):

Action:	**Press CL TRAIN switch**	***Start CL training***
Voice Direct 364:	"Say word one"	*Training CL word*
User:	"Sound System"	
Voice Direct 364:	"Repeat"	
User:	"Sound System"	
Action:	**Press TRAIN switch**	***Start SD training***
Voice Direct 364:	"Say word one"	*Training first SD word in the set*
User:	"Start"	
Voice Direct 364:	"Repeat"	
User:	"Start"	
Action:	**Press TRAIN switch**	***Continue SD training***
Voice Direct 364:	"Say word two"	*Training second SD word in the set*
User:	"Shutdown"	
Voice Direct 364:	"Repeat"	
User:	"Shutdown"	
Action:	**Press TRAIN switch**	***Continue SD training***
Voice Direct 364:	"Say word three"	*Training third SD word in the set*
User:	"Lights"	
Voice Direct 364:	"Repeat"	

Multi-Word CL Mode

The module should be configured for Multi-Word CL Mode (Figur 3). The training sequence is very similar to Single Word CL mode except that 3 separate CL words can be trained. After training eac CL command, up to 5 SD words can be trained. ***It is important t note that each SD set must be trained immediately followin the corresponding CL word. Once the next CL word i trained, there is no way to go back and re-train an SD se without erasing the entire memory.*** Trying to train 2 similar C words will result in a "too similar" error. During SD training, eac SD word is checked for similarity to other words in its set, but no against those in other sets.

Multi-Word CL mode can be useful for a single user product usin multiple gateway commands, or for multi-user applications wher each user would train his or her own CL and SD words.

Example (Multi-Word CL):

Action:	**Press CL TRAIN switch**	***Start CL training for User 1***
Voice Direct 364:	"Say word one"	*Training first CL word*
User:	"Sound System"	
Voice Direct 364:	"Repeat"	
User:	"Sound System"	
Action:	**Press TRAIN switch**	***Start SD training for User 1***
Voice Direct 364:	"Say word one-one"	*Training first SD word in first set*
User:	"Start"	
Voice Direct 364:	"Repeat"	
User:	"Start"	
Action:	**Press TRAIN switch**	***Continue SD training for User 1***
Voice Direct 364:	"Say word one-two"	*Training second SD word in first set*
User:	"Shutdown"	
Voice Direct 364:	"Repeat"	
User:	"Shutdown"	
Action:	**Press CL TRAIN switch**	***Start CL training for User 2***
Voice Direct 364:	"Say word one"	*Training second CL word*
User:	"Sound System"	
Voice Direct 364:	"Repeat"	
User:	"Sound System"	

Example (Multi-Word CL) continued:

Action:	**Press TRAIN switch**	***Start SD training for User 2***
Voice Direct 364:	"Say word two-one"	*Training first SD word in second set*
User:	"Start"	
Voice Direct 364:	"Repeat"	
User:	"Start"	
Action:	**Press TRAIN switch**	***Continue SD training for User 2***
Voice Direct 364:	"Say word two-two"	*Training second SD word in second set*
User:	"Shutdown"	
Voice Direct 364:	"Repeat"	
User:	"Shutdown"	

6. Recognizing Words

SD Mode

When the -RECOG pin is pulled to GND for at least 100mS, recognition will begin. VOICE DIRECT™ 364 will prompt "Say a word". If the response is not recognized, VOICE DIRECT™ 364 will say "Word not recognized" and exit recognition mode. If the spoken word matches a stored template, one (or two) of 8 output pins is activated (pulled high for approximately one second) and a voice message indicates the matching response. If the set contains 8 or fewer elements, these pins may be used to control actions directly. If the set contains more than 8 elements, decoding is necessary. The logical format of the outputs is shown in the table below.

Recognition Word	OUT 1	OUT 2	OUT 3	OUT 4	OUT 5	OUT 6	OUT 7	OUT 8
SD Word 01	A							
SD Word 02		A						
SD Word 03			A					
SD Word 04				A				
SD Word 05					A			
SD Word 06						A		
SD Word 07							A	
SD Word 08								A
SD Word 09	A							A
SD Word 10		A						A
SD Word 11			A					A
SD Word 12				A				A
SD Word 13					A			A
SD Word 14						A		A
SD Word 15							A	A

Note: "A" indicates that the outputs are "Active-high".

If an error occurs during recognition (except for "Word not recognized"), then the error will be spoken "Spoke too soon", "Please talk louder", etc., and the ERROR line will pulse high for 1 second. If the spoken word is not recognized for any reason, none of the pins is activated and an appropriate voice message is synthesized. Since this can occur if VOICE DIRECT™ 364 has not been trained on the word, the "word not recognized" output is not considered an error condition, and the ERROR bit remains inactive.

For each SD word, VOICE DIRECT™ 364 monitors the background noise level and gives a voice warning if the noise is too high for recognition. VOICE DIRECT™ 364 works well with high (above 80 dB) levels of *steady* background noise (such as a fan), but it may make errors at lower levels if the background noise is not steady (for example, a TV set). Best results will be obtained in a relatively quiet location.

Warnings may also appear if the word is spoken too softly, too loudly, or too quickly after the prompt. The ERROR signal goes high when these conditions occur.

Single Word CL Mode

Upon power-up or reset, VOICE DIRECT™ 364 checks to see if it has been configured in a Continuous Listening mode. If any words have been trained, VOICE DIRECT™ 364 will start listening for the CL word. After training, the RECOGNITION button can be pressed to initiate continuous listening operation.

It is important to note that the module is only listening when the TALK LED is lit. Use the LED as an indicator of when you can speak commands, especially when first using the module. After successful recognition of the CL word, there is a 3 second window during which VOICE DIRECT™ 364 is listening for an SD word (if any are trained). If no SD words are trained, then a successful recognition of the CL word will cause the outputs to behave as if the first SD word was recognized (OUT1 would toggle high). This feature is useful if only a single command is required, such as a light switch. Note that VOICE DIRECT™ 364 does not need to wait the full 3 seconds if it hears your command once it detects the silence at the end of your utterance, recognition processing will begin immediately.

When a trained sequence of words (CL word + SD word) is recognized, the appropriate Output Pin(s) will pulse high for 1 second:

Recognition Word	OUT 1	OUT 2	OUT 3	OUT 4	OUT 5	OUT 6	OUT 7	OUT 8
CL + SD Word 01	A							
CL + SD Word 02		A						
CL + SD Word 03			A					
CL + SD Word 04				A				
CL + SD Word 05					A			
CL + SD Word 06						A		
CL + SD Word 07							A	
CL + SD Word 08								A
CL + SD Word 09	A							A
CL + SD Word 10		A						A
CL + SD Word 11			A					A
CL + SD Word 12				A				A
CL + SD Word 13					A			A
CL + SD Word 14						A		A
CL + SD Word 15							A	A

Multi-Word CL Mode

Multi-Word CL mode works in a similar fashion to Single Word mode, except that the module is continuously monitoring for up to 3 words at the same time.

The word (SD) following the triggering word (CL) will be matched only against the templates in the corresponding set. If no SD words are trained, then a successful recognition of one of the CL words will cause the outputs to behave as if the first SD word in that set was recognized. For example, if no SD words were trained for Set B, then OUT2 and OUT4 would toggle high. This feature is useful in applications requiring up to 3 commands, or where there are multiple users.

CL Word	SD Word	OUT 1	OUT 2	OUT 3	OUT 4	OUT 5	OUT 6	OUT 7	OUT 8
CL Word A	SD Word 01	A			A				
CL Word A	SD Word 02	A				A			
CL Word A	SD Word 03	A					A		
CL Word A	SD Word 04	A						A	
CL Word A	SD Word 05	A							A
CL Word B	SD Word 01		A		A				
CL Word B	SD Word 02		A			A			
CL Word B	SD Word 03		A				A		
CL Word B	SD Word 04		A					A	
CL Word B	SD Word 05		A						A
CL Word C	SD Word 01			A	A				
CL Word C	SD Word 02			A		A			
CL Word C	SD Word 03			A			A		
CL Word C	SD Word 04			A				A	
CL Word C	SD Word 05			A					A

Note: "A" indicates that the outputs are "Active-high".

7. Erasing the Templates

When both the –TRAIN and –RECOG pin are pulled to GND for at least 100mS, VOICE DIRECT™ 364 will erase all trained templates. VOICE DIRECT™ will say "Memory erased", pulse the ERROR line high for 1S, and then exit the erase mode. Individual words or sets cannot be erased separately.

8. Tips for Optimal Performance

Successful recognition begins with the careful selection of words for each recognition set. Several factors contribute to selecting an optimal recognition set. Problematic recognition sets can often be corrected by replacing one or more words with a synonym, or approximate synonym (see examples, below), without requiring any other changes. The smaller the set, the higher the recognition rate.

The optimal set consists of:	Avoid sets like:	Aim for sets like:
Dissimilar sounding words	hat/cat/rat home phone/office phone	hat/kitten/mouse home/office
Varying numbers of syllables	orange/apple/cherry	orange/watermelon/ grape

Key considerations for successful voice recording

- The equipment used to train the voice recordings should match the equipment used during recognition. Differences in microphone, microphone housing, etc. will adversely affect recognition.
- The conditions and environment in which the voice recordings are made should reflect the conditions and environment in which the end product will be used.

a) **Distance.** The distance of the microphone from the speaker's mouth must be the same during recording and during end-product use. For example, a doll is typically held within arm's length, and so the voice recording microphone should be held accordingly.

b) **Natural Voices.** Subjects should speak in their normal voice and should be discouraged from sounding different by imitating a foreign accent or using any unnatural intonation. They should be prompted by means of some non-verbal source (pictures or flashcards, for example), so as not to unconsciously mimic the voice of the person doing the prompting.

c) **Physical States.** Physical states should be considered. For example, in collecting voice recordings for an exercise machine, it is strongly recommended to record people who are out of breath.

d) **Emotional States.** Emotional states should be considered. Will the end users be relatively quiet and calm (say, for an office product) or loud and excited (say, for a playground toy)?

e) **Environment / Background Noise.** Environmental noise must be considered. Voice recordings should ideally be made in an environment similar to the one in which the end product will be used. For a product intended for use in cars, it is advisable to record in a running, moving car. Nevertheless, the recordings must be made in a reasonably quiet environment. The speech signal must be prominent relative to background noise and there should not be any abrupt, loud noises. Voice recordings should not be made in a soundproof room. These rooms lend an unnatural background silence to the recordings, which does not reflect the real-world environment in which the end product will be used.

9. About Speech Recognition

VOICE DIRECT™ 364 performs speaker-dependent discrete word recognition by comparing a pattern generated in real time with previously trained word templates. The pattern generated by VOICE DIRECT™ 364 is based on a digital reconstruction of the voice command. Each word to be recognized must first be *trained.* During training, VOICE DIRECT™ 364 constructs a *template* representing the individual speaker's unique sound pattern for each specific word or phrase to be recognized. Templates are stored in serial EEPROM memory. During recognition a new pattern is produced and compared to the stored templates to determine which word was spoken. VOICE DIRECT™ 364 features integrated speech prompting for both training and recognition operations, allowing the development of sophisticated interactive products with minimal programming. VOICE DIRECT™ 364 performs the following operations when recognizing a word:

1. The audio signal (spoken word) is externally amplified and filtered and then supplied to the analog inputs of the VOICE DIRECT™ 364, which converts the analog waveforms to digital samples.

2. VOICE DIRECT™ 364 analyzes the speech signal samples and generates a pattern of information representing significant speech elements.

3. VOICE DIRECT™ 364 increases or decreases the gain of the external amplifier as needed to maintain signal quality.

4. Using a neural network, the pattern is compared with previously stored template patterns; a small number of candidate templates is selected.

5. The candidate templates are further processed to determine the one template that provides the best match to the unknown pattern.

6. If the best match template gives a score above a pre-defined threshold, VOICE DIRECT™ 364 chooses the word associated with that template. If no template provides a match above threshold, a special "no match" value is chosen.

Steps 1-3 above are repeated for each word during training. VOICE DIRECT™ 364 stores the average of two training patterns for each word to improve accuracy. Before storing a new template it is compared with the existing templates in the set. The new candidate word will not be accepted if it is too similar to an existing word (for example, "Bill Smith" and "Jill Smith").

VOICE DIRECT™ 364, like other speech recognition systems, is necessarily subject to two types of errors: *rejects* (failure to recognize a word in the vocabulary) and *substitutions* (confusion of two vocabulary words, or recognition of a non-vocabulary word). The relative importance of each type of error may depend on the application. VOICE DIRECT™ 364 provides selectivity levels that allow the user to optimize the tradeoff between these two types of recognition errors. When the recognition selectivity level is set to its highest value, VOICE DIRECT™ 364 minimizes substitution errors but may produce more rejection errors. When the training selectivity level is set to its highest value, VOICE DIRECT™ 364 minimizes both substitution and rejection errors by rejecting words that are too similar sounding, potentially increasing recognition accuracy. These settings are best established by experimenting with your own application and environment.